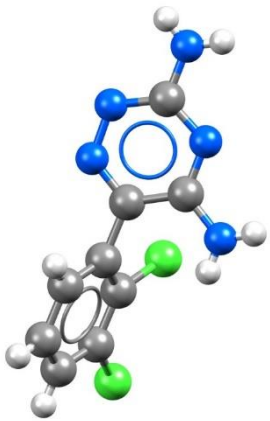
Docking with Water in the Binding Site using GOLD

2021.2 CSD Release



CCDC
advancing structural science

Table of Contents

Docking with Water in the Binding Site 2
- Objectives 3
- Pre-required skills 3
- Materials 3

Case Study 4
- Introduction 4
- Provided Input Files 5

Preparing Protein for Docking Experiment 6

Exploring Ligand for Docking Experiment 7
- Defining the Binding Site 7
- Selecting Ligand for Docking 8

Setting Water in the Binding Site 9
- Running GOLD Docking with water5 turned off 11
- Running GOLD Docking with water5 toggled 13

Analysis of Results 14
- Water5: Off 14
- Water5: toggled 15

Conclusions 16
- Next steps 16

Glossary 17

Docking with Water in the Binding Site

The importance of water molecules found in the *binding site* of a protein lies in their ability to mediate the interactions between the *ligand* and the protein and to form hydrogen bonded networks that can stabilise a protein-ligand complex in solution (See Figure 1). Such a hydrogen-bonded network of water molecules may stabilise the complex formed with one ligand but not another, thus contributing to the specificity of ligand recognition. Additionally, water molecules may also help to stabilise the conformation of the active sites of enzymes. All these mechanisms have been shown to be of importance to drug discovery.[1]

GOLD allows one to perform *docking* experiments accounting for these "active" water molecules. In GOLD, the active water molecules can be retained and allowed to switch on and off (i.e., bound, or displaced) and to rotate around their three principal axes (to optimise hydrogen bonding) during docking.

To predict whether a specific water molecule should be bound or displaced, GOLD estimates the free-energy change associated with transferring a water molecule from the bulk solvent to its binding site in a protein-ligand complex. A constant penalty is added for each water molecule that is switched on and represents the loss of rigid-body entropy on binding the target, hence rewarding water displacement. Therefore, for a water molecule to be bound to a protein-ligand complex, its intrinsic binding affinity needs to outweigh the loss of rigid-body entropy on binding (see GOLD User Guide for further details).

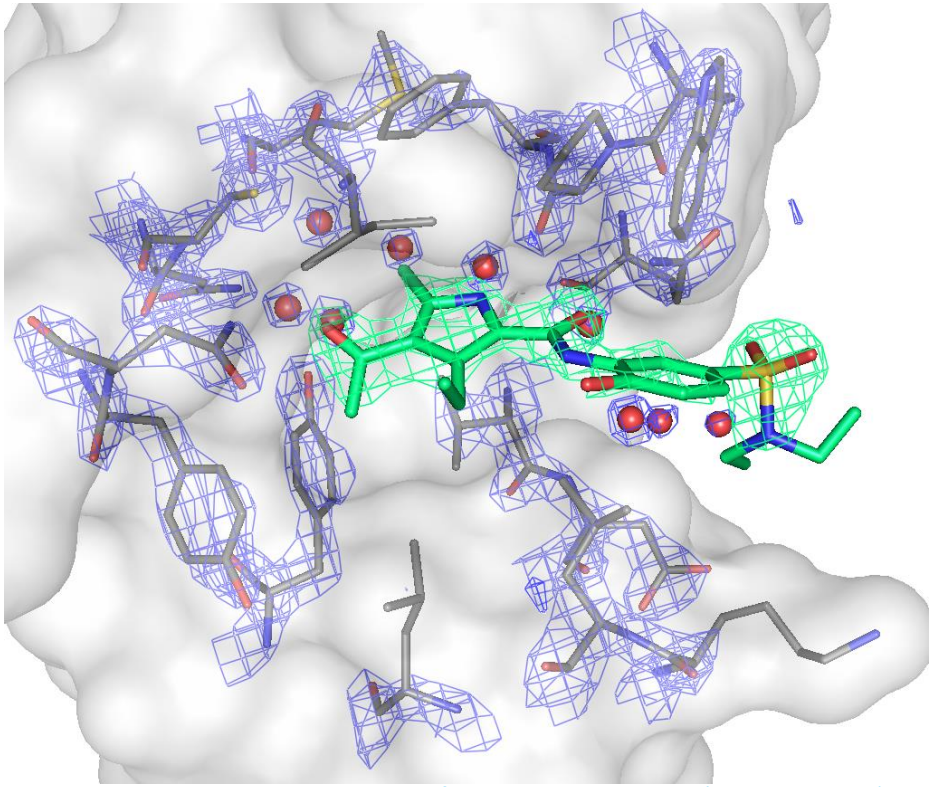


Figure 1. Crystallographic structure of BRD4-XD14 complex (PDB code:4lyw). The electron density map of the binding site contoured at 2.0σ includes water molecules in the binding site. The protein residues are coloured in grey, the ligand in green and the water molecules in red.

[1] Ladbury, J. E. "Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design." *Chem. Biol*. (1996). **3**, 973-980.

Objectives

In this workshop, you will investigate docking to a binding site that contains water molecules which a ligand may either displace, or alternatively, make use of through hydrogen bond interactions. In this workshop, you will:

- Learn how to set up a docking run using the GOLD wizard in Hermes
- Analyse docking results

This workshop will take approximately ***30 minutes*** to complete.

Note: The words in *Blue Italic* in the text are reported in the Glossary at the end of this handout.

Pre-required skills

This example assumes you are already familiar with how to setup protein(s) and ligands for docking calculation. If not, please refer to the following sections of the GOLD User Guide:

- Setting Up the Protein
- Essential Steps

This workshop requires an intermediate user experience level of Hermes and GOLD. We recommend working through the introduction to Hermes workshop that can be found here and looking through the GOLD user guide.

Materials

For this workshop we will use the files that you can download from here.

Case Study

Introduction

Lysine acetylation is a key mechanism that regulates chromatin structure; it is a site-specific modification that directs both structural changes to chromatin and gene transcription. The recognition of acetylated lysine residues by specific "reader" protein domains such as bromodomains (BRDs), has a critical impact in the regulation of gene expression. The role of bromodomains in translating a deregulated cell acetylome into disease phenotypes was recently unveiled by the development of small molecule bromodomain inhibitors. This breakthrough discovery highlighted bromodomain-containing proteins as key players in cancer biology, as well as inflammation and remyelination in multiple sclerosis.[2]

Your protein:

BRD4, like others bromodomains, is characterised by a bundle of four α-helices interconnected by three loop regions of different length. The large amount of co-crystallised structures of BRD4-inhibitors have provided structural knowledge about the binding sites of BRDs. The three key areas are: the acetyl-lysine recognition site, the vacant lipophilic shelf (WPF shelf) and the ZA loop region (see Figure 2).

Your ligand:

XD14 (4-acetyl-N-[5-(diethylsulfamoyl)-2-hydroxyphenyl]- 3-ethyl-5-methyl-1H-pyrrole-2-carboxamide) is a potent inhibitor of BRD4. The co-crystallised structure of BRD4-XD14 (PDB code: 4lyw) reveals that XD14 binds in the acetyl-lysine recognition site mimicking the acetyl-lysine residue interaction with BRD4 (see Figure 3). Positioning the 4-acyl substitution in the pyrrole ring towards the highly conserved Asn140, XD14 engages in a direct hydrogen bond interaction.

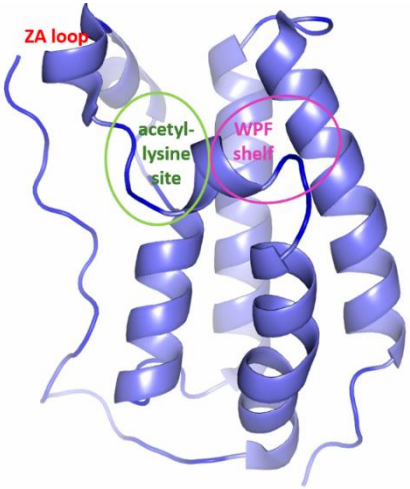


Figure 2. Secondary structure representation of BRD4. The key areas – acetyl-lysine site, WPL shelf and ZA loop – are highlighted in green, magenta and red, respectively.

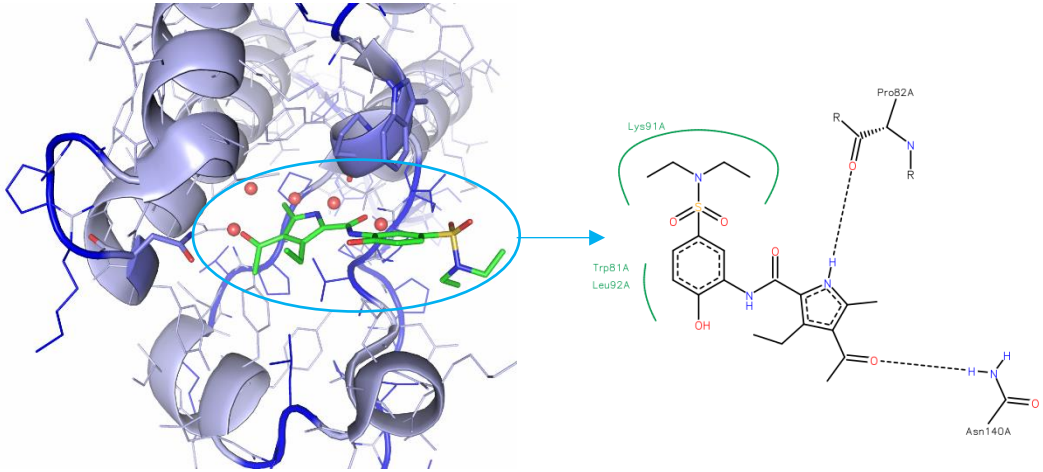


Figure 3. Co-crystallised BRD4-XD14 (PDB code: 4lyw). The protein is in light blue cartoon representation, the ligand in green and the water network is coloured in red. Additionally, we show the 2D interactions diagram of the ligand in complex with BRD4.

[2]Lucas X *et al.* "4-Acyl Pyrrole: Mimicking Acetylated Lysines in Histone Code Reading. *Angew. Chem. Int. Ed.* (2013). **52**, 14055-14059

The pyrrole ring is located deep inside the recognition pocket, and complements the hydrophobic pocket defined by four conserved waters with a 5-methyl substitution. In the co-crystallised structure of BRD4-XD14, an additional water molecule is found to be directly involved in an H-bond with XD14.

Challenges:

Crystallographic structures of BRD4 in complex with different inhibitors have shown that four water molecules are highly conserved in the bromodomains binding site. Interestingly, the binding site of BRD4 in complex with XD14 inhibitor accommodates an additional water molecule that is directly involved in an H-bond with XD14 (see Figure 4).

- In this tutorial, we will evaluate if the addition of the fifth water molecule is required for a more accurate reproduction of the crystallographic complex BRD4-XD14.

The example used here mimics the situation where a researcher has a crystal structure of a protein binding site and is unsure which and how many of the waters in the binding site should be included in the model for use in an inhibitor design effort.

Provided Input Files

- The *protein.mol2* will be used to guide you through the key steps required to prepare a protein for docking experiment.
- The reference ligand file *(XD14_ligand.mol2)* has been extracted from the co-crystal structure.
- The ligand file (*XD14_idealised.sdf*) has been set up in accordance with the guidelines for the preparation of input files (Setting Up the Protein(s) and Setting Up Ligands).
- The active waters (i.e. those that you would like GOLD to consider during docking) are provided in separate files (*water_1.mol2*, *water_2.mol2, water_3.mol2, water_4.mol2 and water_5.mol2).* The original PDB file *4lyw.pdb* is also available, should you wish to prepare protein, ligand and waters from scratch. The active waters used in this study correspond to *HOH1107, HOH1114, HOH1119, HOH1130* and *HOH1140* water molecules in the PDB file.

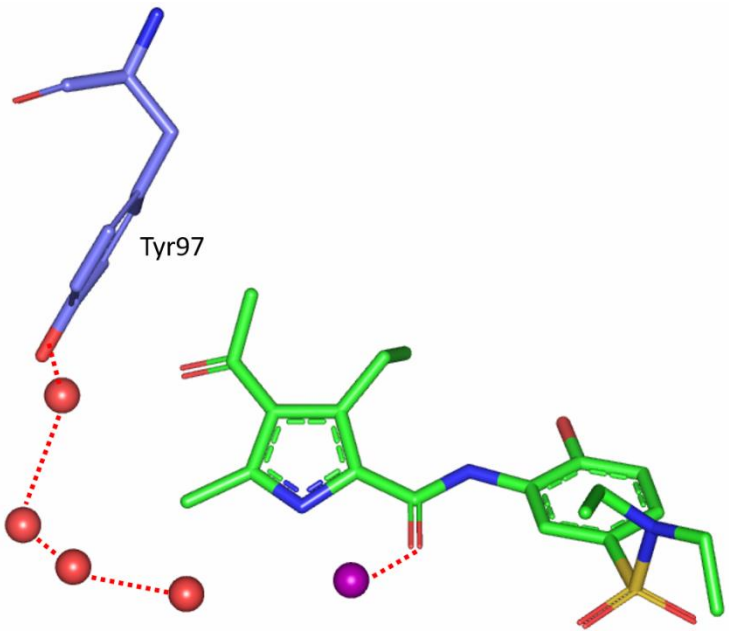


Figure 4. Water network in the binding site of BRD4 complexed with XD14. Four conserved water molecules are represented as red spheres whereas, an additional water molecule specific to this complex and interacting via a H-bond with XD14 is shown as a purple sphere.

Preparing Protein for Docking Experiment

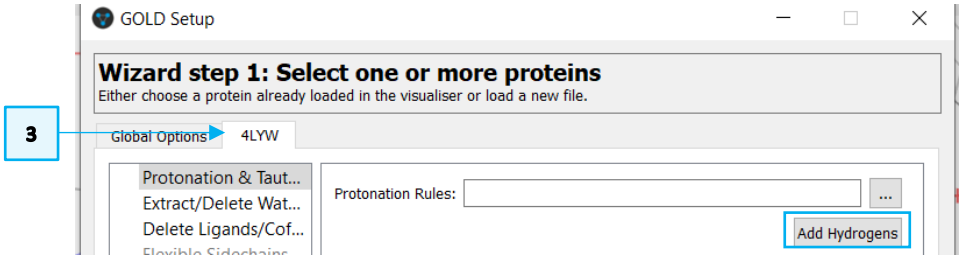
1. Launch Hermes by clicking its icon and click **GOLD**, then pick **Wizard...** from the resultant pull-down menu.

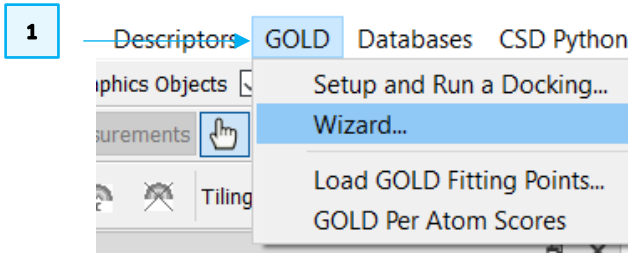
2. Click **Load Protein** in the *GOLD Setup* window and read in the protein file provided with this tutorial (*protein.mol2)*. As the protein is loaded, a tab corresponding to that protein appears to the right of the *Global Options* tab and it is labelled with the name taken from the protein file, *4LYW.*

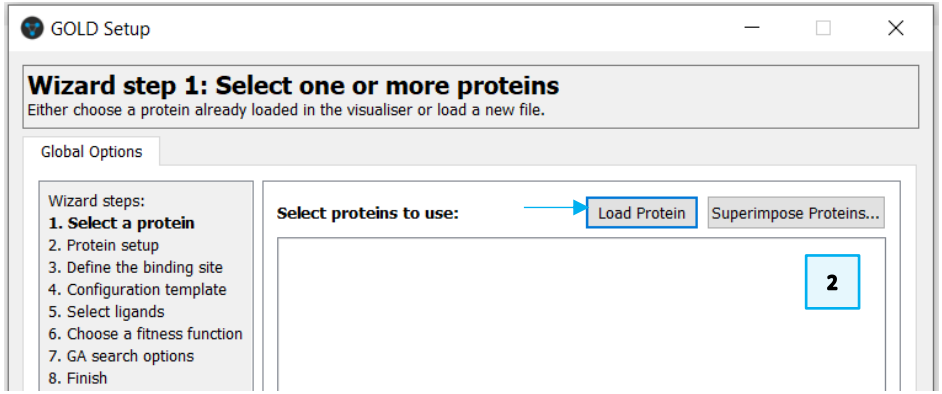
3. Click on the *4LYW* tab adjacent to the *Global Options* tab. From within the *4LYW* tab, add hydrogen atoms to the protein by clicking on **Add Hydrogens** from the *Protonation & Tautomers* option in the Wizard. A pop up window will inform you that 1271 hydrogens were added to the protein.

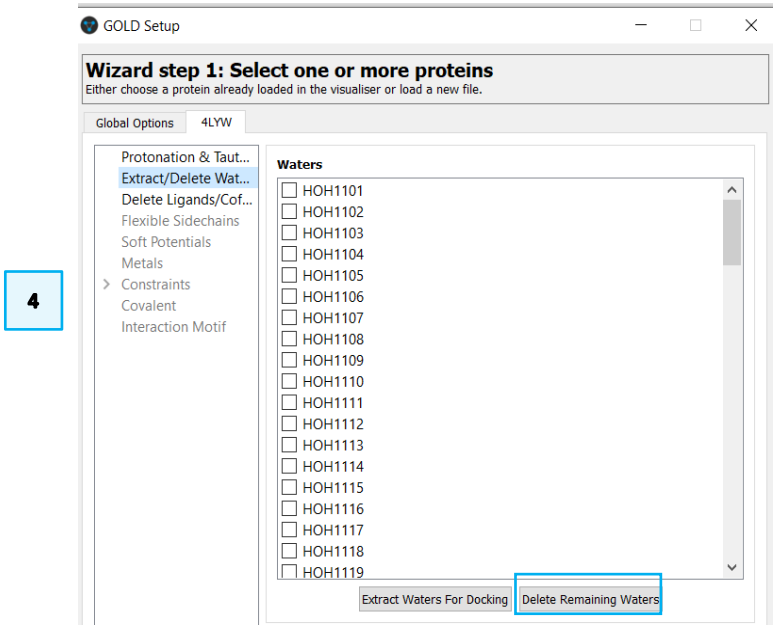
4. Still in *4LYW* tab, click on **Extract/Delete Waters**. From within this dialogue, it is possible to specify some active water molecules and to delete those that are not required. Since the active water molecules have been prepared separately and provided with the input files, it is appropriate to delete all water molecules by clicking on the **Delete Remaining Waters** button. When prompted with a pop-up message saying *Are you sure you want to delete all waters?* click **OK.** You will be informed that 98 water molecules have been deleted. Then click **Next** which returns you to the Protein setup step in the Global Options tab.

5. The protein is now set-up for docking; click **Next** to move to the following step of the Wizard where you can *Define the binding site*.









Exploring Ligand for Docking Experiment

Defining the Binding Site

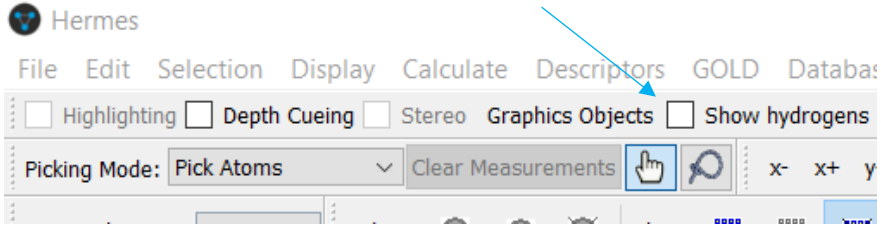
Defining the binding site for docking is a crucial step and can be done in several ways (see GOLD User Guide for further details).

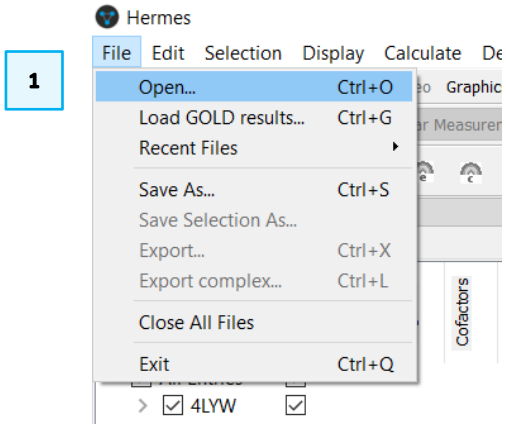
In this tutorial, we will define the binding from a reference ligand.

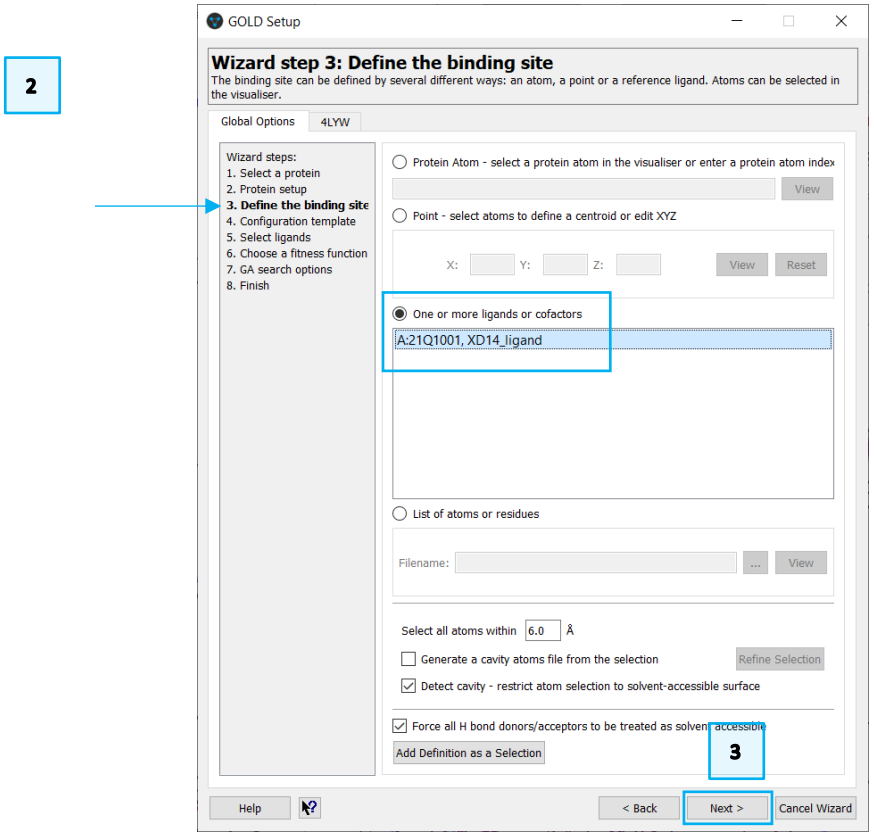
1. Read in *XD14_ligand.mol2* in Hermes by clicking on **File** from the main menu and then **Open** from the resultant pull-down menu. This loads the ligand in the 3D view and makes the ligand available for binding site definition.

2. While on option *3. Define the binding site* from the GOLD Wizard, click the radio button next to **One or more ligands** and click to select *A21Q1001, XD14_ligand* to determine the binding site. Carbon atoms outside of the binding site will turn purple. Leave all other parameters to their defaults. For visualisation purposes, it can help here to switch off the display H-atoms using the **Show hydrogens** tick box in the top-level menu of Hermes. Click **Next** to proceed.

3. At this point you are given the option to load a configuration file template. These templates can be used to load recommended settings for a number of different types of docking protocols (see GOLD User Guide). In this example, we will specify all docking settings manually. Click **Next** to proceed to the *Select ligands* step in GOLD wizard.







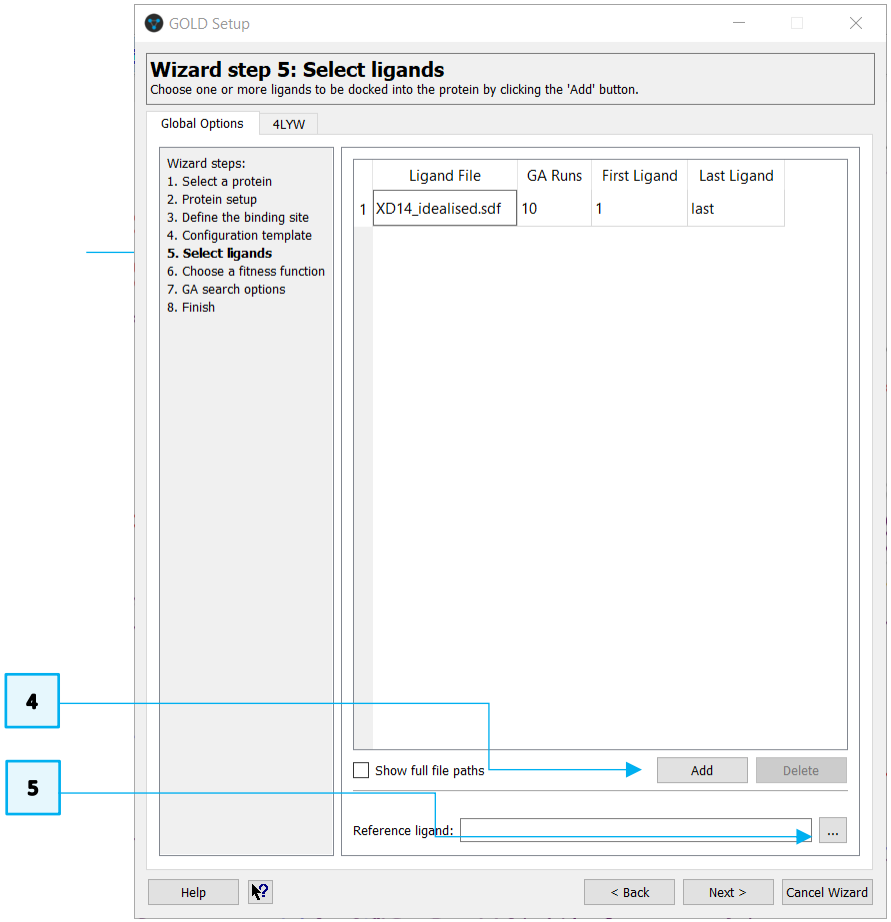
Selecting Ligand for Docking

To proceed with the GOLD setup, we need to specify the ligand that we want to dock into the protein.

4. The prepared ligand is available in the tutorial folder (*XD14_idealised.sdf*). As with the protein file, all hydrogen atoms have been included in this ligand file. Specify this ligand by clicking **Add** at the bottom of the *GOLD Setup* window. The *XD14_idealised.sdf* is now listed under *Ligand File.*

Here you have the option to supply the file containing the reference ligand (i.e. the ligand in the native *conformation*) that will be used to perform an automated RMSD calculation with respect to the reference ligand conformation. For each GOLD solution, the resultant RMSD value will be written out to the files containing the fitness function ranking, i.e. the ligand rank file (.rnk) and bestranking.lst file.

5. To specify the ligand reference file either enter the path and the filename of the file in the **Reference ligand** box, or click on the **...** button and use the file selection window to choose *XD14_ligand.mol2* file that is provided in the tutorial folder. This is the native conformation of XD14 in complex with BRD4 protein.



Setting Water in the Binding Site

Before proceeding with the docking setup, we must define the active waters that we want to be considered during the ensemble docking. Five active water molecules have been detected, four of which define the hydrophobic pocket that locates the pyrrole ring of XD14 and a fifth water molecule which is directly involved in hydrogen bond with XD14. The active waters must be provided in separate files, one water molecule per file and can be found in the tutorial folder.

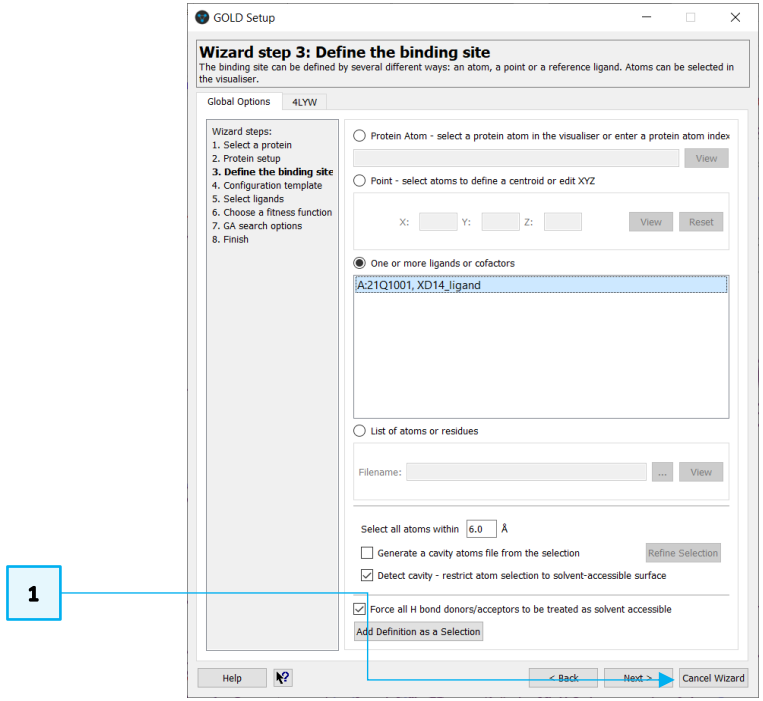
1. *GOLD Wizard* provides the key steps for docking; however, more advanced options (i.e., *Configure Waters)* are available outside the main Wizard. To access these advanced options, click **Cancel Wizard**.

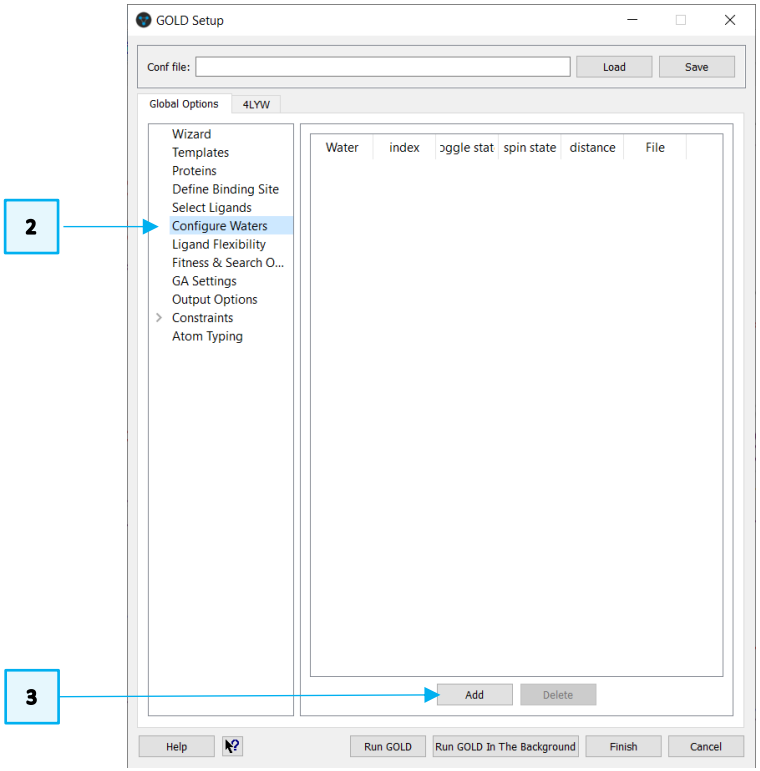
2. Pick **Configure Waters** from the list of available options.

3. The dialogue is empty, so we need to specify our water molecules by reading in the water files. To do this click on the **Add** button, select the five water files then click **Open**.

4. The water molecules will be listed in the *Configure Waters* dialogue. By default, each water molecule in the *Configure Waters* list will be retained in the binding site during docking and will be allowed to spin and toggle to optimise the position of the molecule and the orientation of the hydrogens. These settings can be customised for specific water molecules within the *Waters* dialogue in the *GOLD Setup* window.

   - *Toggle* state leaves GOLD to decide whether the water should be present or absent (bound or displaced by the ligand) during the docking.
   - *On* sets the water to be always present in the binding site and allows the hydrogen positions to vary during docking in order to maximise the hydrogen bonding score both from interactions with the protein and the ligand.





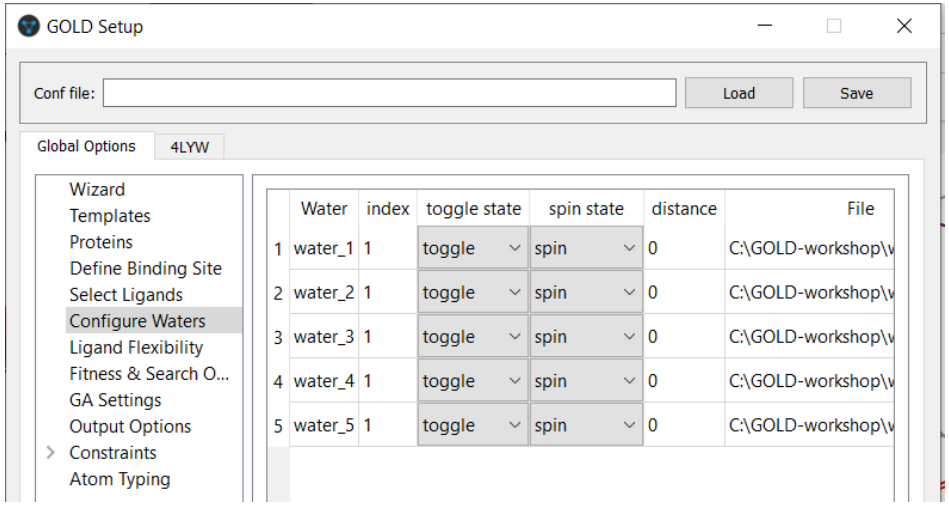
- The *Off* water state option allows a water to be removed from consideration during docking.

The orientation of the waters can be also changed.

- Activating the *spin* option makes GOLD automatically optimise the orientation of the hydrogen atoms.
- Activating the *trans_spin* option and inserting a translation value into the distance dialogue makes GOLD spin and translate the water molecule to optimise the orientation of the hydrogen atoms as well as the water molecule's position within a defined radius. Note that the distance value must be between 0 and 2 Å.
- Activating the *fix* option makes GOLD to use the orientation specified in the input file.

Because in this tutorial we want to evaluate if the fifth water molecule contributes to a better prediction of the experimental binding pose, four conserved water molecules (water_1, water_2, water_3 and water_4) will be always considered in the binding site, whereas the presence and displacement of the water_5 will be evaluated.

4



Running GOLD Docking with water5 turned off

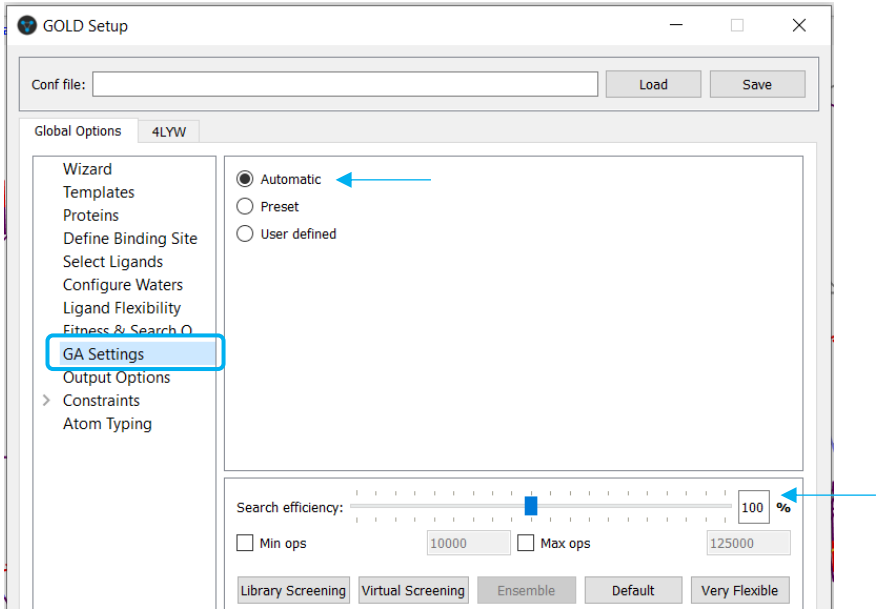
5. In the *Configure Waters* dialogue of GOLD Setup change the *toggle state* of the first four water molecules to **on**. Then, change the *toggle state* of the water_5 to **off**. This means that we will see how XD14_idealised is accommodated in the binding site in absence of the fifth water molecule.

6. Click on *Fitness & Search Options* to continue with the GOLD Setup. Change the ***Scoring Function*** to be GoldScore by choosing this option from the dropdown menu.

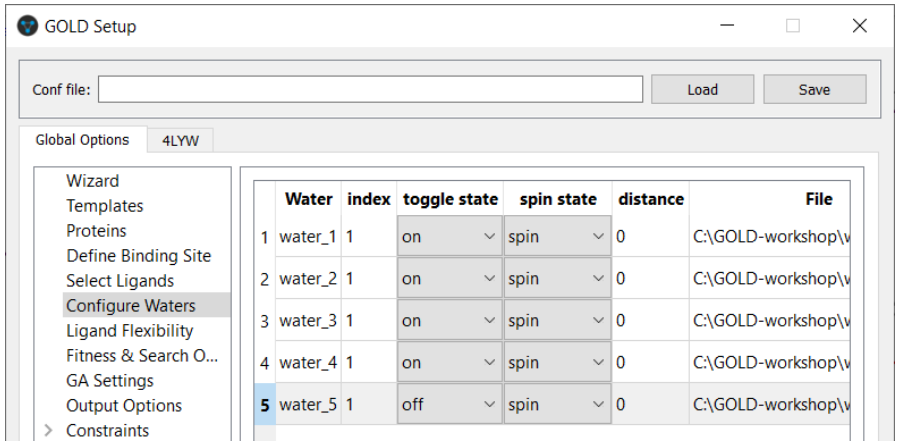
7. By default, the **Allow early termination** check box is ticked. Deactivate this by unticking the box. This will ensure that as many solutions as possible are explored.

8. Click on the *GA Settings* dialogue and enable automatic *GA setting* by clicking **Automatic** radio button; ensure the *Search efficiency* is set to 100%. This will make GOLD automatically calculate an optimal number of operations for a given ligand, thereby making the most efficient use of the search time.

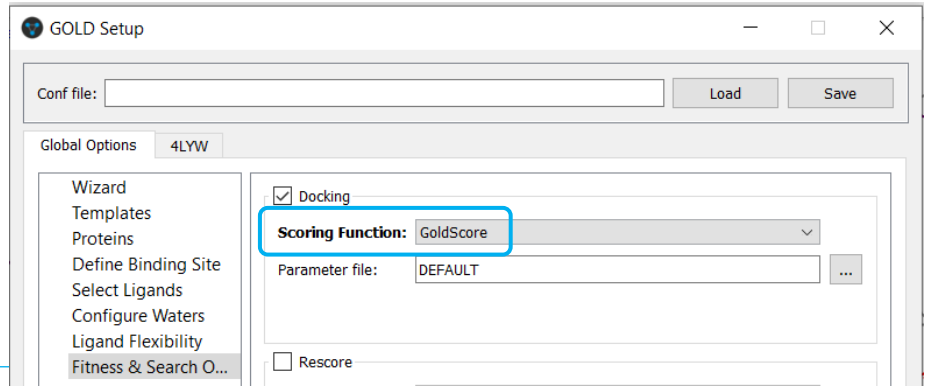
8



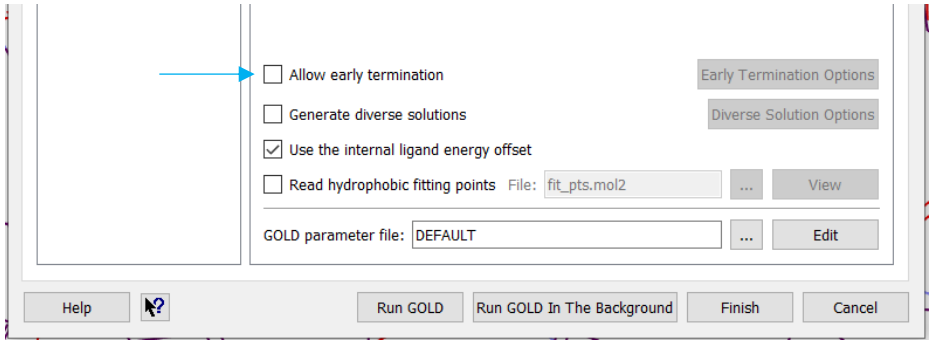
5



6



7



9. Before starting the docking, select *Output Options.* Click the **…** button next to *Output directory* and either browse to or create an appropriate directory, *e.g. water5_off.*

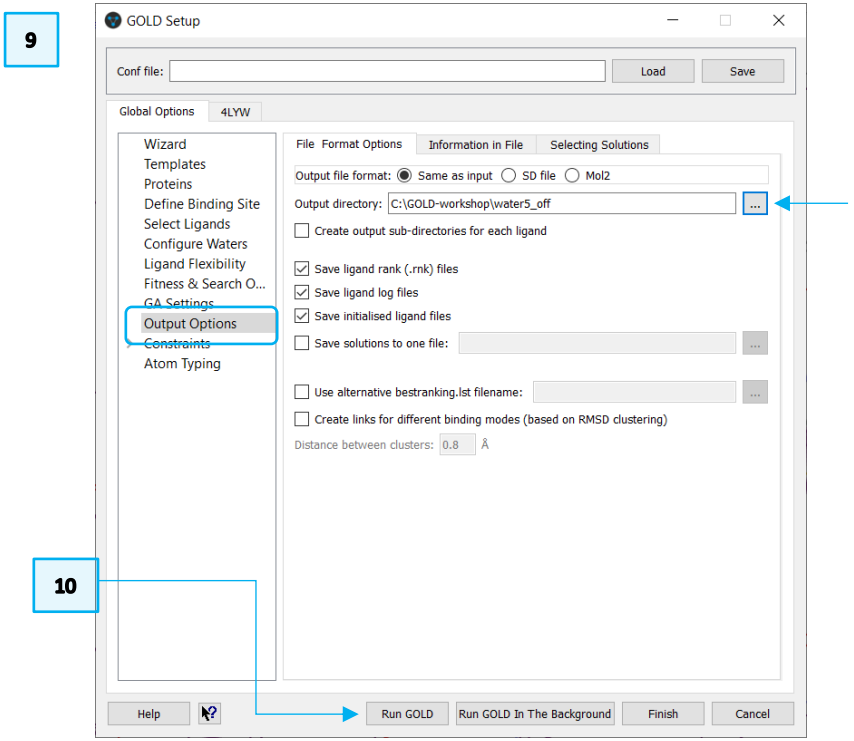
10. We have now finished setting up our docking, so click on the **Run GOLD** button at the bottom of the GOLD interface. You will be presented with a *Finish GOLD Configuration* window containing *Save Files* options.

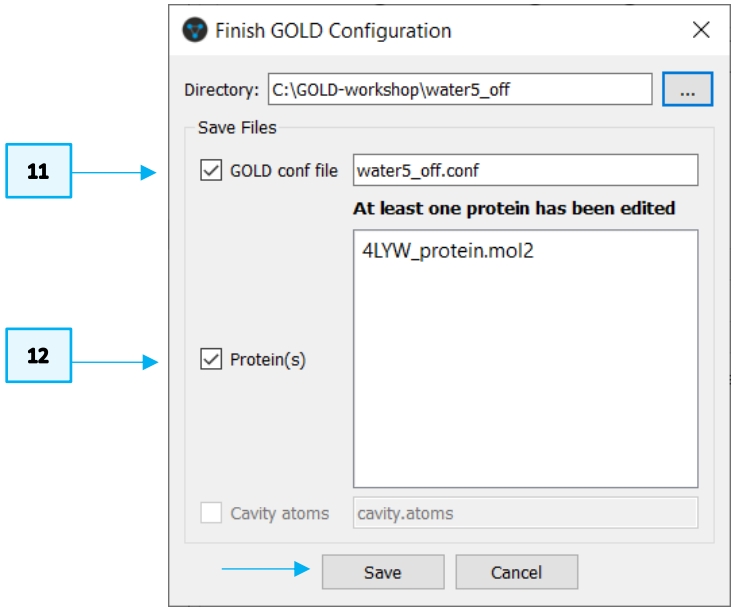
11. Ensure the *GOLD conf file* box is ticked and type in the filename box to rename the conf file as *water5_off.conf.*

12. Ensure that Protein(s) box is ticked. To save the edited 4LYW structure, click **Save** to start the docking.

13. As the job progresses output will be displayed in several tabs in the *Run GOLD* window. Note that the docking experiment will take several minutes.

14. Once the job is complete, the message *Finished GOLD batch job, exiting…* will appear in the *gold.log* tab of the *Run GOLD* window. Additionally, the **View Solutions** button in *Run GOLD* window will become available.





Running GOLD Docking with water5 toggled

15. Return to the GOLD front end by clicking **Close** in the *Run Gold* dialogue box. The button will become active once the run is completed.

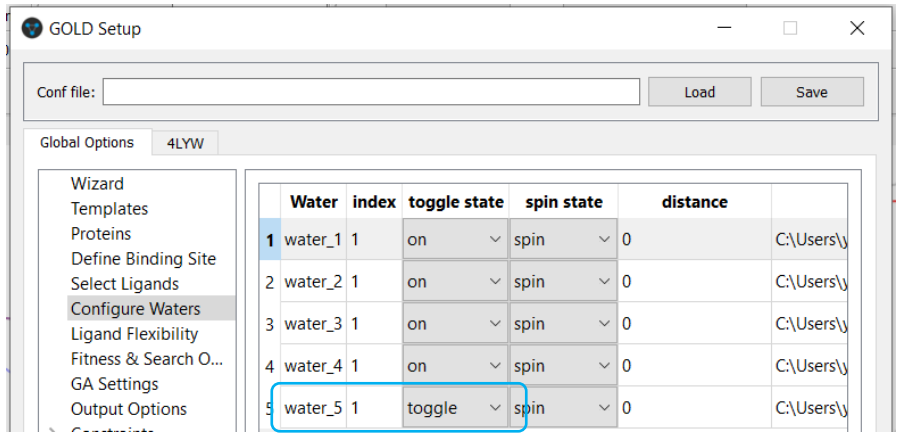
16. Click on **Configure Waters** to bring up the water setup window. Change the *toggle_state* of the *water_5* molecule to **toggle**. This means that we leave GOLD to decide whether the *water_5* should be present or absent during the docking and that the waters orientation of the hydrogen atoms will be optimised for all the water molecules.

17. Click on **Output Options** and change the output subdirectory name *e.g. water5_toggle.* There is no need to change any other settings.

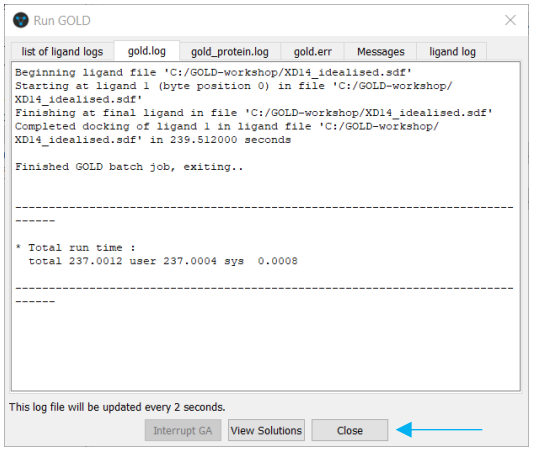
18. Click **Run GOLD**. In the *Finish GOLD Configuration* window, as before, edit the name of the GOLD configuration file in the GOLD conf file text box to *e.g. water5_toggle.conf*. This time the Protein(s) tick box is off because we did not edit the protein. Click **Save**.

19. Once the GOLD run has finished, click **Close** in the *Run GOLD* window, **Cancel** in the *Gold Setup* window, and tidy up the 3D view by clicking the top-level **File** menu then selecting **Close All Files**.

16

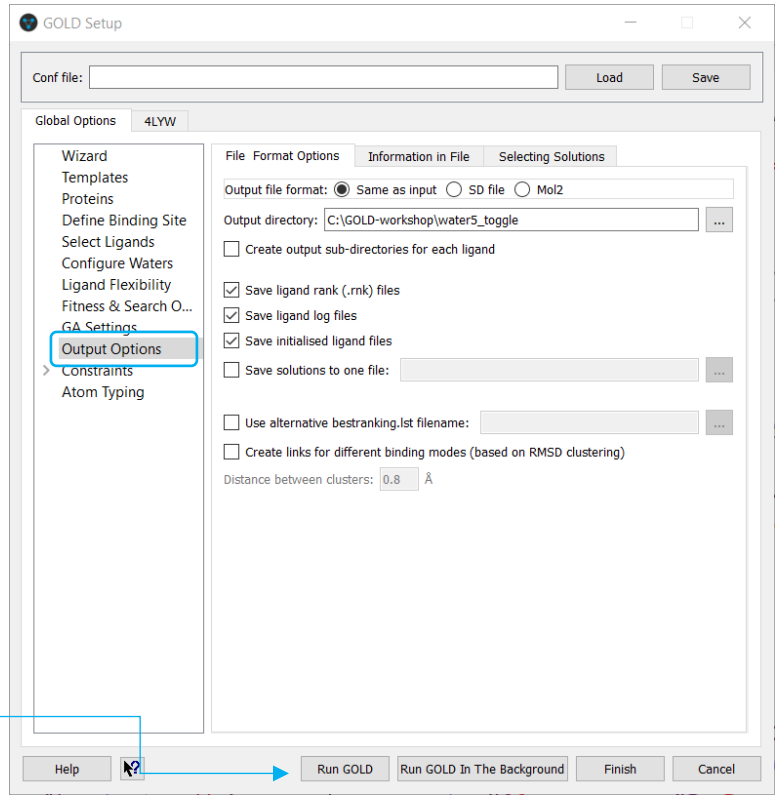


15



17

18



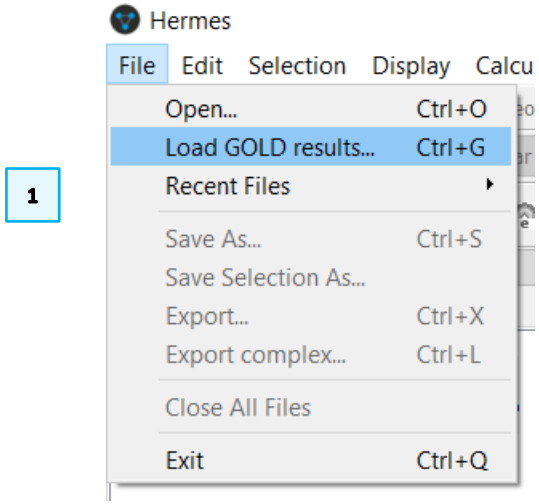
Analysis of Results

We will now evaluate how the different settings of the fifth water molecule will affect the docking of XD14 to the BDR4 protein.

Note: In the following examples, you may wish to modify the representation style or the colour of the reference ligand so that it is clearer. Once the files are loaded, click on the *Display* tab in the *Molecule Explorer*, then right-click on the reference ligand. From the dropdown menu, it is possible to modify the style and colour of the reference ligand. You can also undisplay the Chain once the result is loaded.

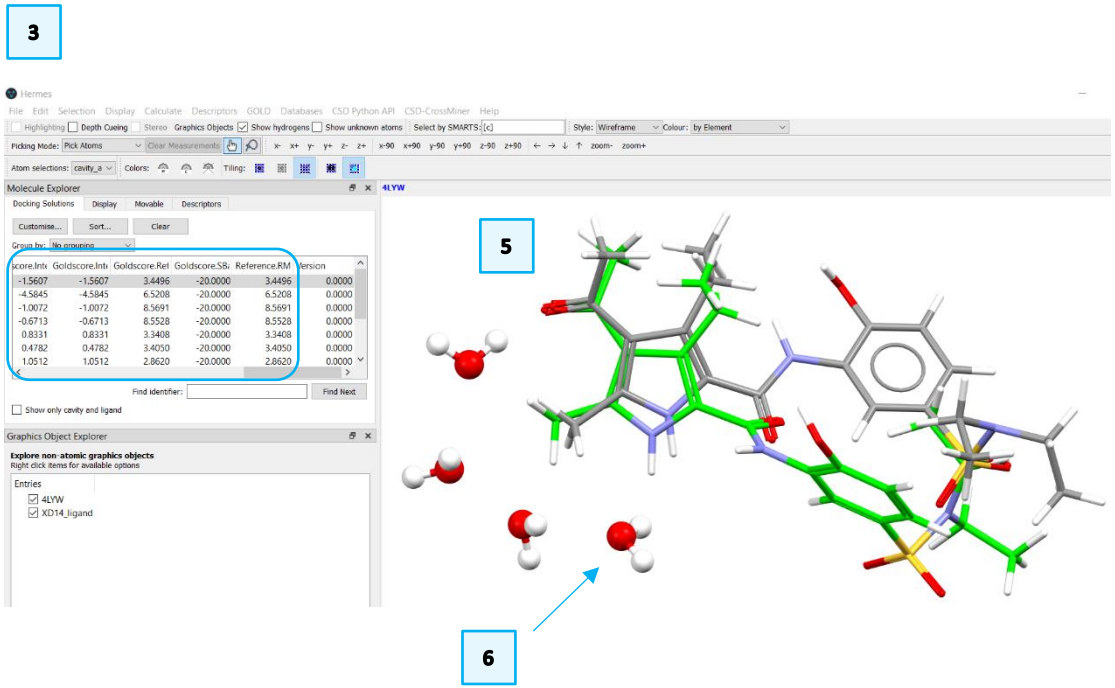
Water5: Off

1. Load the results of the *water5_off* docking by clicking **File** > **Load GOLD results** then select the *water5_off.conf* file.



2. Load the reference ligand by clicking on **File**, **Open**, then navigate to the tutorial directory which contains the *XD14_ligand.mol2* file. Select the file then click **Open**. Optionally you can also load the five water molecules.

3. The docking solutions are given in their docked order with their corresponding fitness score under the column headed *Goldscore.Fitness*. If desired, the solutions can be ordered by clicking on this *Goldscore.Fitness* header to determine which is the highest scoring. We have obtained 10 docking solutions as this is the default number of how many times our ligand was docked in the BRD4 protein.



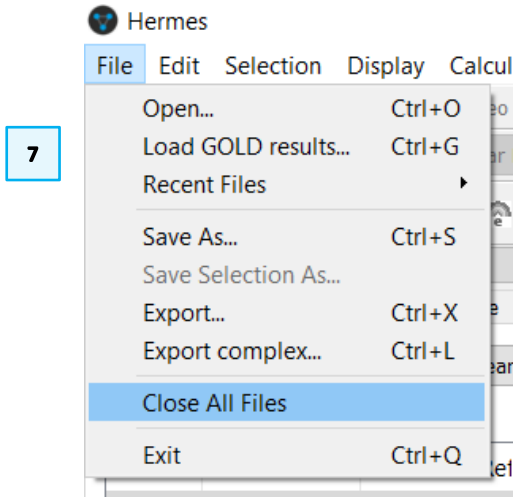
4. Scroll through the docking solutions to check their poses against that of the reference ligand. You should find that several docking solutions are found, none of which closely resemble the correct binding mode.

5. The RMSD of the top ranked pose when compared to that of the co-crystallised ligand is 3.45 Å. (Note that docking results and RMSD values may

vary due to the non-deterministic nature of GOLD). The top-ranking pose matches the orientation of the pyrrole ring; however, it fails to match the orientation of the diethyl sulphonamide and hydroxyphenyl groups. Here, our reference ligand is coloured by Element and displayed as Capped Sticks.

6. By inspecting the binding mode, we can see that the hydrogen positions of the four water molecules were optimised to form the water network characteristic of the BRD4 binding site.

7. From Hermes click on **File** and select **Close all files** from the pull-down menu to clear the 3D-viewer of Hermes.

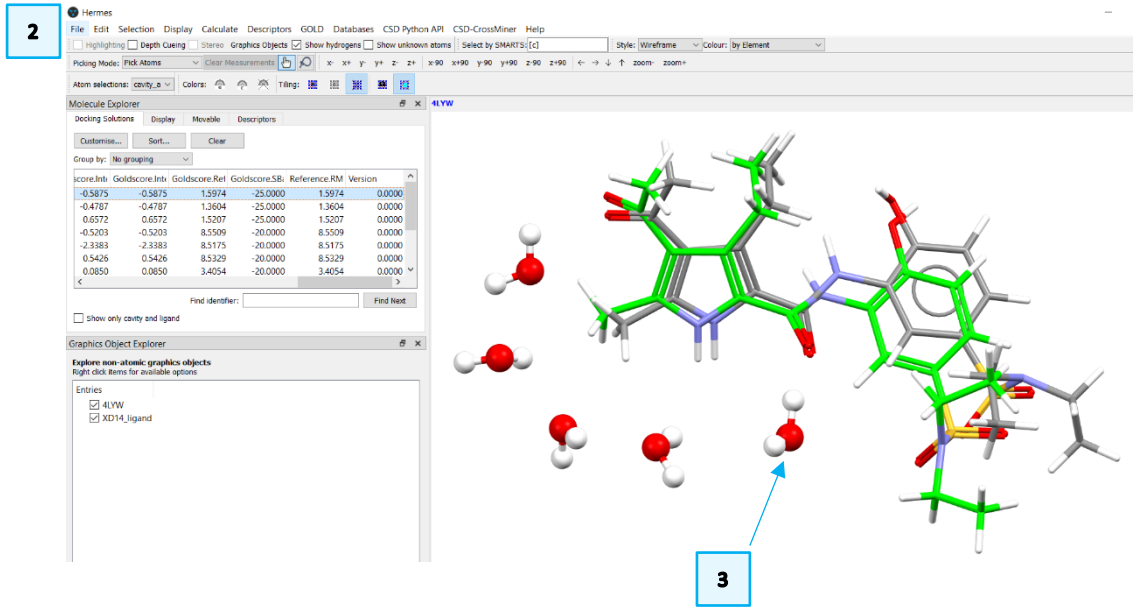


Water5: toggled

1. From **Load GOLD results** in Hermes, load the *water5_toggle.conf* file and load the reference *XD14_ligand.mol2* file as in steps 1 and 2 above.

2. Check each solution and compare the different poses to that of the reference ligand. The top-ranking pose here shows a binding pose similar to reference ligand with an RMSD of 1.60 Å.

3. By inspecting the binding mode, we can see that the fifth water molecule has been included in the best ranked poses.



Conclusions

- When only four water molecules are considered in the binding site, none of the solutions replicate the original binding mode. The pyrrole ring is correctly placed; however, the size of the protein binding site allows the diethyl sulphonamide and hydroxyphenyl groups of XD14 to occupy an alternative area of the cavity than in the original crystal structure.

- When the fifth water molecule is considered, the native binding mode can be replicated.

- The additional water molecule is required to correctly orient XD14 in the binding site through a direct H-bond that coordinate the ligand with the water network constituted by the other conserved four water molecules (Figure 3).

Next steps

After this workshop, you can continue learning about GOLD with more exercises available in the self-guided workshops available in the CSD-Discovery workshops area on our website.
https://www.ccdc.cam.ac.uk/Community/educationalresources/workshop-materials/csd-discovery-workshops/

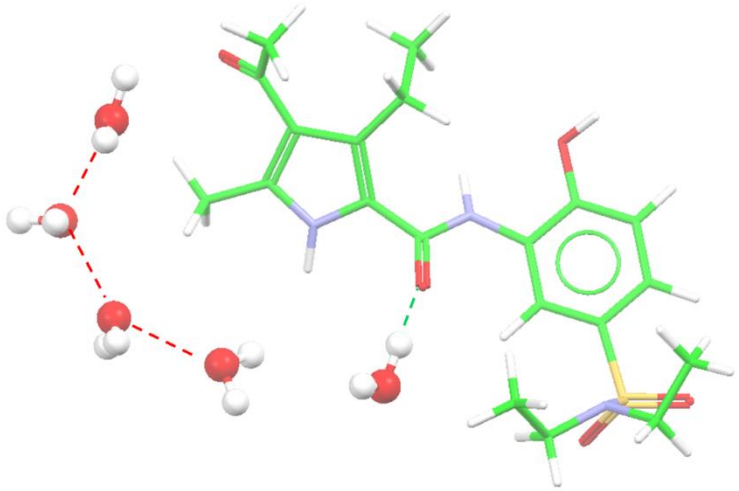


Figure 3. Binding mode of the top-ranking pose. H-bonds within the water network are coloured in red and H-bond between XD14 and the additional water molecule is coloured in green.

Glossary

Binding Site

A specific region (or atom) in a molecular entity that is capable of entering into a stabilizing interaction with another molecular entity. An example of such an interaction is that of an active site in an enzyme with its substrate. Typical forms of interaction are by hydrogen bonding, coordination and ion pair formation. Two binding sites in different molecular entities are said to be complementary if their interaction is stabilizing. *Source*: PAC, 1994, 66, 1077. (Glossary of terms used in physical organic chemistry (IUPAC Recommendations 1994)) on page 1089.

Conformation

The spatial arrangement of the atoms affording distinction between stereoisomers which can be interconverted by rotations about formally single bonds. Some authorities extend the term to include inversion at trigonal pyramidal centres and other polytopal rearrangements. *Sources*: PAC, 1994, 66, 1077. (Glossary of terms used in physical organic chemistry (IUPAC Recommendations 1994)) on page 1099.

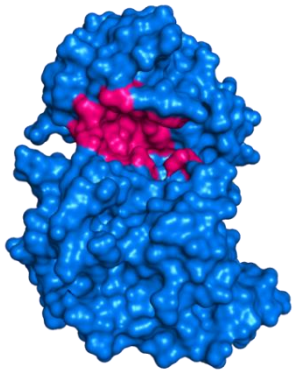
Docking

Docking studies are computational techniques for the exploration of the possible binding modes of a substrate to a given receptor, enzyme or other binding site. *Source*: PAC, 1997, 69, 1137. (Glossary of terms used in computational drug design (IUPAC Recommendations 1997)) on page 1142.

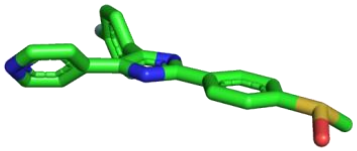
GA setting

A genetic algorithm is an optimization algorithm based on the mechanisms of Darwinian evolution which uses random mutation, crossover and selection procedures to breed better models or solutions from an originally random starting population or sample. *Source*: PAC, 1997, 69, 1137. (Glossary of terms used in computational drug design (IUPAC Recommendations 1997)) on page 1144

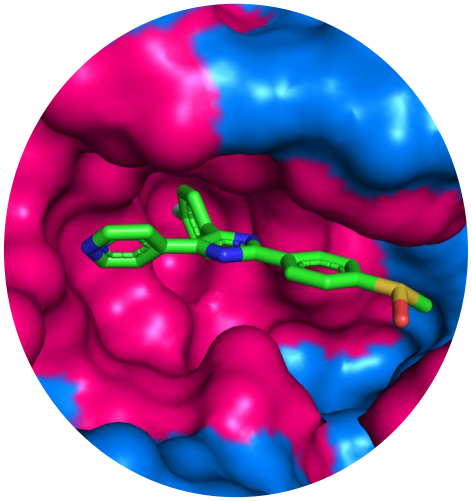
In GOLD, number of genetic operations performed (crossover, migration, mutation) is the key parameter in determining how long a GOLD run will take (i.e.



A protein binding site highlighted in pink



A ligand



Docking calculates a pose for the ligand and a score for the pose

this parameter controls the coverage of the search space). GOLD can automatically calculate an optimal number of operations for a given ligand, thereby making the most efficient use of search time, e.g. small ligands containing only one or two rotatable bonds will generally require fewer genetic operations than larger, highly flexible ligands. The criteria used by GOLD to determine the optimal GA parameter settings for a given ligand include:

· the number of rotatable bonds in the ligand,

· ligand flexibility, i.e. number of flexible ring corners, flippable nitrogens, etc. (see Ligand Flexibility),

· the volume of the protein binding site, and

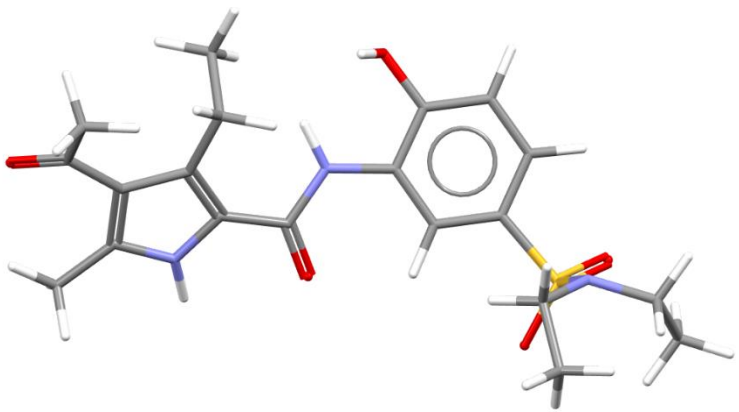
· the number of water molecules considered during docking (see Water Molecules).

The exact number of GA operations contributed, e.g. for each rotatable bond in the ligand, are defined in the gold.params file (see Altering GOLD Parameters: the gold.params File). For further information please refer the GOLD user manual.

GOLD
GOLD (Genetic Optimisation for Ligand Docking) is a genetic algorithm for docking flexible ligands into protein binding sites. GOLD has been extensively tested and has shown excellent performance for pose prediction and good results for virtual screening.

Ligand
In this context, these are the molecules under investigation. These could be organic, inorganic and macromolecules.

Scoring function
In the fields of computational chemistry and molecular modelling, scoring functions are mathematical functions used to approximately predict the binding affinity between two molecules after they have been docked. Most commonly one of the molecules is a small organic compound such as a drug and the second is the drug's biological target such as a protein receptor.



XD14 (4-acetyl-N-[5-(diethylsulfamoyl)-2-hydroxyphenyl]- 3-ethyl-5-methyl-1H-pyrrole-2-carboxamide) is a potent inhibitor of BRD4 and ligand docked in this workshop